DTU Library

Human Error Probabilites (HEPs) for generic tasks and Performance Shaping Factors (PSFs) selected for railway operations

Thommesen, Jacob; Andersen, Henning Boje

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Human Error Probabilites (HEPs) for generic tasks and Performance Shaping Factors (PSFs) selected for railway operations

Report 3.2012

DTU Management Engineering

Jacob Thommesen
Henning Boje Andersen
Februar 2012

DTU Management Engineering
Department of Management Engineering

Title:
Human Error Probabilities (HEPs) for generic tasks and Performance Shaping Factors (PSFs) selected for railway operations

Authors:
Jacob Thommesen
Henning Boje Andersen

Summary:
This report describes an HRA (Human Reliability Assessment) of six generic tasks and four Performance Shaping Factors (PSFs) targeted at railway operations commissioned by Banedanmark. The selection and characterization of generic tasks and PSFs are elaborated by DTU Management in close collaboration with Banedanmark. The estimates provided are based on HRA literature and primarily the HEART method, being recently been adapted for railway tasks by the British Rail Safety and Standards Board (RSSB). The method presented in this report differs from the RSSB tool by supporting an analysis at task level, which can be performed with fewer resources than a more detailed analysis of specific errors for each task. The generic tasks are presented with estimated Human Error Probabilities (HEPs) based on and extrapolated from the HRA literature, and estimates are compared with samples of measures from comparable tasks from the COREDATA database. PSFs are presented with multipliers to be used in combination with generic tasks types to support a quantitative HRA of railway tasks. Estimates contained in this report should be used with caution and judgment, since they are largely based on estimates derived from industries other than rail and the general warning that a task-based analysis is less precise than an error-based one. The authors recommend that estimates be adjusted to actual measures of task failures when feasible.

Version	**Date**	**Remarks**
Draft 1	19.12.2011	Preliminary lists of tasks and PSFs
Draft 2	12.01.2012	Definitions of agreed tasks and PSFs. Samples from COREDATA
1.0	03.02.2012	Extrapolated estimates for HEPs and PSF multipliers
2.0	08.02.2012	Interim copy to Banedanmark
2.01	06.02.2012	Minor revisions. Accepted by Banedanmark
2.02	15.03.2012	Minor revisions. Final Issue. Accepted by Banedanmark

ISBN: 978-87-92706-40-9

Department of Management Engineering
Technical University of Denmark
Produktionstorvet 424
2800 Kongens Lyngby
Denmark

Contents

Intentionally left blank

Introduction

The work described in this report was commissioned by Banedanmark (BDK) in October 2011 and has been prepared November 2011-January 2012 by the authors. Banedanmark has proposed and contributed to the selection of generic tasks and performance shaping factors; Banedanmark has also provided examples of concrete tasks that are described in the report and are intended to serve as instances of the respective generic tasks under which they are listed. Finally, Banedanmark has provided essential input to the scoping of report and indispensable feedback to early drafts.

The generic tasks that are listed and analyzed in this report have been selected to represent tasks that are, first, applicable to the rail domain and, second, specifically representative of operator tasks following the changes to Danish railways operations upon adoption of the Signalling Programme. The report is therefore intended to support BDK within the Signalling Programme in hazard analysis and in the preparation and critique of specific safety levels arguments as part of the design and of procedures and associated skills and training requirements of the new signalling systems.

There are few empirical data (historic records or simulation-based) or expert judgment data available that exactly matches the generic tasks provided. Therefore, the estimates provided in this report are based on extrapolation from the error rates provided in the HRA literature, primarily from HEART. For each estimate, sources and the extrapolation are described.

The choice of a first-generation method such as HEART might perhaps be questioned in the light of advances in HRA methods and techniques since the appearance of HEART more than 25 years ago. First-generation methods focus on the skill and rule based level of human action and are often criticized for failing to consider the impact of context, safety culture and other organizational factors and, not least, so-called errors of commission. Such errors arise typically when an operator misjudges (wrong diagnosis, faulty situation awareness) a complex, abnormal situation and then, with the best intentions, performs acts that increase risks or perhaps even produce an accident. Techniques such as HEART are designed to capture and quantify the risk of failing to perform required actions and are not well suitable to predict and capture actions performed by operators outside the scope of predefined or predicted task scenarios. But for a new and yet to be deployed application such as the Signalling Programme it is appropriate to begin by applying a first-generation method, whereas the more complex application of methods that may capture lack of resilience against more rare and ill-defined situations may be better addressed when training with skills and knowledge requirements have been defined.

For each of the generic task estimates provided, we have performed an independent appraisal of the estimates by selecting from the COREDATA database (Gibson et al. 2006;Williams 1986) a set of estimates of tasks that, as far as possible, match the generic tasks The appraisal should not be regarded as a proper validation, nor has it been possible in all cases to identify a reasonably large set of tasks (n>20) in the COREDATA database that exemplify the generic task in question.

Note on using HEART for estimates

The choice of HEART as the primary source of error rates is based on the consideration that this technique incorporates the most widely used estimates of error rates of *generic* tasks; moreover, HEART is the most prominent technique used by the RSSB in their studies of HRA applications (Hickling 2007). Finally, a chief advantage of HEART is that it provides a *task*-based analysis with a ‘holistic’ approach (Kirwan 1994) rather than a ‘decompositional’ approach focusing on types of *error* which would require a higher level of detail.

There are, however, some challenges when using HEART. One challenge is that definitions of GTTs (Generic Task Types) are sometimes very specific and complex. In particular, they often include performance shaping factors (PSFs), which may lead to a risk of ‘double counting’, i.e., including a PSF in the definition of a generic task and subsequently finding a need for using this specific PSF to modify the task for a specific application (Cullen et al. 2004:41;Hickling 2007). The risk of double counting can be somewhat controlled by either defining restrictions on the use of PSFs for specific GTTs – and such matrices are already available (Hickling 2007:66-7) – or by extracting PSFs from GTT definitions & estimates.

Reversibility and checking

Another challenge is that HEART estimates seem pessimistic (when compared with UK incident data). Hickling has suggested that HEART estimates may be too conservative for railway driving tasks (Hickling 2007:60), partly because they do not account for (1) performance improving features such as 'independent checking' that are already included in the actual task architecture, or for (2) "*considerable opportunity to recover from an initial error*".

The first of these reliability enhancing factors – checking – is partly included in one of the PSFs presented in this report: 'supervision' (see later for further explanation). We shall thus assume that the generic tasks in this report do not include supervision and checking, and that such features will require application of the PSF.

The benefits of supervision will also depend on the second factor, opportunity to recover, or **task reversibility**, since a supervisor would otherwise have limited opportunity to check and correct. We shall – to some extent – follow Hickling's assumptions: that HEART estimates do not account for reversibility, and that most tasks covered by BDK's analysis are in fact reversible.

Hickling argues that "*the error probability given in HEART, and the application of HEART in most risk models, represents a single opportunity for a task error before an undesired consequence emerges. This is appropriate in the context of process control as most assessed task will consist of tasks comprising a number of small sequentially executed actions that are often irrecoverable if they are incorrect*" (Hickling 2007:61). Hickling's observation about assumed irreversibility in HEART estimates should to some extent be counterbalanced by the fact that HEART actually has a PSF for 'irreversibility' (EPC 7, *8, see table A.2), implying that generic tasks are basically reversible – with task B as the exception, as explained below and in William's corrective matrix (Hickling 2007). We shall nevertheless accept Hickling's argument about the need to correct (reduce) HEART estimates to account for reversibility, although with less than the full multiplier presented in HEART.

Furthermore, we assume that tasks to be analyzed by BDK are basically reversible, unless specifically and explicitly noted to be irreversible. Hickling argues that this is the case for train-driving tasks: "*the driver has multiple opportunities to respond correctly to a signal, but it is only at the last point where there is insufficient time for their action to be effective that the error becomes critical*" (Hickling 2007:61). Some might question whether this is also true for signalling tasks, as argued by an English supervisor: "*essentially once a signaller has pressed a button to clear a signal there is little that can be done to rectify a regulation error*" (Dickinson & Lowe 2007:227). However, BDK argues that problems of irreversibility are be reduced with modern signalling systems like CBTC and ERTMS level 2 as continuous radio connection to the onboard equipment introduces the options of shortening of movement authorities as well as the remote application of unconditional emergency stops. Further mitigating action can be achieved by use of GSM-R voice Railway Emergency Call that automatically connects a controller voice call to a group of drivers within an area – with the option of instructing the driver to command their trains to a standstill. It is therefore reasonable to regard the generic tasks described in this report as basically reversible – and therefore more receptive to the benefits of supervision.

Multiplier identified from HEART

As we have observed, one way of taking account of pessimistic estimates in HEART may be to modify HEART estimates to account for checking and opportunity to recover. Hickling suggests identifying a modifier based on existing HEART estimates by comparing two complementary GTTs in HEART (Table A.1):

B ('without supervision or procedures', 'single attempt', HEP=0.26), and

F ('following procedures, with some checking', HEP=0.003).

Based on the definitions, the difference between B and F seems to represent three positive factors:

1) procedures,
2) supervision (incl. checking)
3) reversibility (as opposed to 'single attempt').

These three factors together account for the difference in HEP estimates by a factor of about 90. (This is rounded up to a multiplier of 100 by Hickling (2007:62)).

In other words: F = B*procedures*supervision*reversibility (where all three factors are positive and reduce error rates, and thus with a value < 1), with the accumulated value: procedures*supervision*reversibility = HEP(F)/HEP(B) = 1/90. All three factors are included in this report: reversibility is assumed in generic tasks, supervision is a PSF, and procedures are presumed to be included in skill-based and rule-based tasks in the rail domain. As will be explained, the PSF 'job aid' will be defined as involving an external device or support (e.g., a checklist represented on a memory card; an alarm when a limit is exceeded) and will thus always involve a procedure.

Basic tasks with modified estimates

Having contrasted Generic tasks B and F, Hickling then identifies Generic task C ('complex task requiring high level of comprehension and skill') as a **'baseline cognitive task'** – and modifies it by 1/100 (an approximation of the modifier discussed above) to get 0.16/100 = 0.0016. He also identifies Generic task E ('Routine, highly-practiced, rapid task involving relatively low level of skill') as a '**baseline skilled task**' (HEP=0.02) – and applies the same multiplier to get HEP=0.0002.

Note on using COREDATA for validation of estimates

We have decided to seek an independent basis for appraising the suggested estimates. To do so, we compare the estimates with data sampled from COREDATA (Gibson et al 2006), using a version of the database provided by Mr. Huw Gibson, RSSB, to the authors (21st November 2011). A proper validation would require a more thorough and rigorous sampling than is possible for the limited time and resources available for the preparation of this report; nevertheless, the samples provided can be seen as an extra check or triangulation against the estimates collected from other sources.

There are, however, some challenges to this comparison. COREDATA presents probabilities for specific *errors* in specific task scenarios, and there may be several possible errors for each task. Some errors occur at different stages of an evolving situations, i.e. with several possibilities for recovery from an initial error. When we focus on the error for a specific task with several opportunities for recovery, we should only count those errors that are not recovered.

In this report, we shall only make a crude comparison, and resources are not available to make a more detailed analysis of the error rates presented in COREDATA. Such an analysis lies outside the scope of the current report, but we shall make some indications about the limitations and possible further analysis.

Parameters in COREDATA used for selection of records

The main data in COREDATA are listed in a table of HEP (Human Error Probability) values for different tasks or failure types, collected from different studies, 413 records in total. The table also has other attributes that are used for selection in this report.

A number of studies used in COREDATA are dedicated to communication – and some communication-related tasks – primarily in Air Traffic Control (studies no. 16, 35-38, 54-56), and one study in railway communication (study no. 34). Records from these studies are selected – using the attribute 'Study ID' – as the basic sample for tasks 3 and 4.

The attribute 'Level_of_operation' has several possible values: 'Emergency', 'Maintenance – Routine', 'Maintenance – type not identified', 'Normal Operation', 'Normal/emergency', 'Perturbed Operation' and 'Various'.

This attribute has been used to identify emergency tasks: 93 records are thus captured as exemplifying ‘Emergency’.

The attribute ‘Data_Pedigree’ indicates the source of estimates for individual studies with text strings such as ‘Simulator Task Monitoring’, ‘Incident Data/Near Miss Data’ or ‘Expert Judgment’. This attribute is used to identify results based on expert judgment.

The attribute ‘Task_Familiarity_I’ can be read as indicating task frequency with values: ‘More than once an hour’, ‘Daily’, ‘Weekly’, ‘Monthly’, ‘Yearly’, ‘Between 2 and 9 years’, ‘Between 10 and 40 years’ and ‘Unlikely during a working lifetime’. For many records (272), however, such data are not available, indicated as ‘Various’ or ‘Data not available’ or empty.

The attribute ‘Procedure_vs_Problemsolve’ is used to rate the degree of problem solving/proceduralisation by values 1 to 5, “where 5 is pure problem solving and 1 is highly proceduralised”. For many records (139), however, this information is not available, indicated as either ‘Various’ or empty. This attribute is used often in this report to distinguish proceduralised tasks from tasks with some degree of problem solving.

Besides the ‘HEP Table’, a subtable is used to indicate ‘Cognitive error modes’ for some tasks/failure types (records in HEP Table), often with more than one error mode for each task/failure. The possible error modes are: ‘Attention’, ‘Decision Making’, ‘Long Term Memory’, ‘Short Term Memory’, ‘Perception’ and ‘Response Execution’.

1. Generic tasks

In collaboration with Banedanmark (BDK) the authors have identified a small number of generic tasks of particular use to BDK. In the table below we provide an overview of these tasks, their definitions, key references and resulting estimates of Human Error Probability (HEP) and, when available, intervals that show the 5-95 % range.

There are seven basic tasks, the first (‘task 0’) being a ‘lower limit’ of human performance.

The next two are basic tasks that do not involve a sense of urgency associated with emergency, and are not communication tasks. One is a basic simple routine task, while generic task 2 is a knowledge-based task requiring judgment.

Tasks 3 and 4 are mediated verbal communication tasks representing two different types of – typically radio-based – communication between a signal man and other operators such as a driver or a track worker in charge of a track possession.

The last two tasks represent different types of emergency tasks that are both characterized by some sense of urgency, which will also involve some level of stress. The urgency is induced either by safety-critical situation or production pressure (without imminent safety aspects). The main distinction between tasks 5 and 6 is based on the degree of urgency as well as the availability of a plan of action and relevant information.

Table 1. Generic tasks

Generic tasks			
Title	**Description**	**HEP**	**References**
0. Human performance limit	Highest obtainable reliability (minimal error rates).	1E-4 (single) 1E-5 (team)	(Kirwan 1994)
1. Simple routine	Simple, familiar and frequent task skill-based or rule-based, for which procedures are available	4E-4 (0.00014-0.0009)	HEART, (Hickling 2007)
2.Nontrivial familiar	Familiar, relatively frequent task, requiring knowledge-based performance and judgment. (Normal tasks involving a deviation from planned operations.)	1.6E-2 (0.012-0.028)	HEART, (Hickling 2007)
3. Communication, routine	Familiar content routinely conveyed, and where at least a limited template for communication is available, and error capture and correction is possible.	6 E-3	(Kirwan et al. 2004;Kirwan & Gibson 2007)
4. Communication, nonroutine	Unforeseen, novel content, where no template – or only a rudimentary one – is available.	3 E-2	(Kirwan 1994)
5. Emergency scenarios - known	Task characterized by some urgency and stress due to safety or production concerns, for which a plan of action (a template, a script) and relevant information are available.	1 E-1	(Kirwan 1994)
6. Emergency scenarios - unknown	Task characterized by high degree of urgency and stress due to safety or production concerns, where no adequate plan of action (a template, a script) is available, and relevant information is uncertain or missing.	3E-1	(Kirwan 1994)

When using the generic tasks to analyze actual tasks in Banedanmark the HEP values are to be modified by performance shaping factors as appropriate to take account of the presence of supervision, time pressure, job aids and/or task competition. Basically, all generic tasks do thus not include these performance factors. In some cases, however, we will discuss possible overlaps between generic task descriptions and PSFs.

Generic task 0: Human performance limit

Definition

Highest obtainable reliability (minimal error rates).

References

Kirwan suggest two different 'human performance limits', one for a single operator and a lower value for teams (see Table A.3 of appendix A):

- '*17. Human performance limit: single operator*' (HEP =1E-4)
- '1*8. Human performance limit: team of operators performing a well-designed task, very good PSFs, etc.*' (HEP=1E-5)

Generic task 1: Simple routine

Definition

Simple, familiar and frequent task performed at the skill-based or rule-based level[1], for which procedures are available.

There is no time pressure assumed here. However, the task may be embedded in a time pressure task scenario, and in this case the increased error rate may be accounted for by the corresponding PSF. Alternatively, if there is some urgency and stress involved in specific task, then it properly belongs to task 5.

Communication tasks are not covered by this task.

References

Two Generic Task Types (GTT) in HEART (see table A.1) seem to cover the most routine tasks:

- G (familiar, frequent, adequate time, without 'job aid'; HEP=4E-4),
- E (Routine, highly practiced, rapid, low skill level); HEP = 2E-2.

Mention of 'rapidity' in the definition of task E seems to imply both Time Pressure (TP) and no or low reversibility, suggesting HEP may be lower for reversible tasks without time pressure (defined as a PSF in this report). In other words: **E(HEART) = task1*TP*irreversibility** (where both factors are negative and thus with a value > 1). Thus, we shall modify (reduce) HEP for task E to account for task 1 characteristics: reversibility and lack of time pressure. Similarly, Hickling suggest modifying task E to get a '**baseline skilled task**' (Hickling 2007:63), although we cannot simply apply Hicklings modifier (0.01), since it represents two other factors besides reversibility: procedures, supervision/checking.

Furthermore, two types of error from Kirwan's generic guideline data (Table A.3) are relevant for this task:

---

[1] The distinctions between skill-based, rule-based and knowledge-based performance are based on Rasmussen's model (Rasmussen 1986) and "*well-known and liked*", and experience shows that they "*support a HEART like quantification phase*" (Hickling 2007:32)

- ’*12. Error of omission of an act embedded in a procedure*’ (HEP=3E-3).
- ‘*14. Error in simple routine operation*’ (HEP=1E-3)

The definition of Kirwans *error* 12 is based on one specific error type (rather than *task* type): there will often be several possible errors for one task, e.g. the task can fail in other ways than omission. One may thus express the full error rate of a corresponding *task* ‘kirwan12’ (‘act embedded in a procedure’): **HEP(kirwan12) = HEP(error12) + x**, where HEP(error12) is the value presented above, and where x represents the probability of other errors for the same task. Furthermore, Kirwan refers to a procedure, which our task1 doesn’t. Thus, if task1 is similar in other characteristics to Kirwan’s *task* 12 then the following applies: **procedure*HEP(task1) = HEP(error12)+x**, where ‘procedure’ represents an improvement and thus a value < 1.

Estimates

We base our estimates on HEART probabilities for the reasons explained in the introduction. We shall follow Hicklings suggestion and use a modification of task E. For this modification a reduced version of Hicklings modifier (see p. **Fejl! Bogmærke er ikke defineret.**) is applied, representing only procedures, reversibility and adequate time, without supervision: 0.02.

HEP (task1) = 4E-4, interval: 0.00014-0.0009.

Rail examples from Banedanmark

- Driver parks train, apply runaway protection
- Driver evaluates train safe and fit for service
- Signalman adds control room log entry
- PICOP using handheld terminal to request pre-planned possession
- PICOP placing dual-faced stop markers indicating the limits of a work site
- Watchman at lookout duty warns of approaching train
- Signalman manually routing trains on interlocked routes
- Track worker passing tracks at staff crossing
- Driver observing trackside markers
- Driver readying for departure, closing doors
- Track worker staying within defined safe workzone
- Signalman performing manual route release for unused route
- Signalman manually throwing point
- O&M Coordinator resets axel counter

Sampled records

Selection criteria

COREDATA contains a total of 413 records. We have excluded:

- all 116 records involving communication failures (from communication studies).
- all 92 records involving emergency tasks are excluded (and reserved for tasks 5 and 6).

These selections result in a set of non-communication and non-emergency records, a total of 205 (413 – 116 - 92) records.

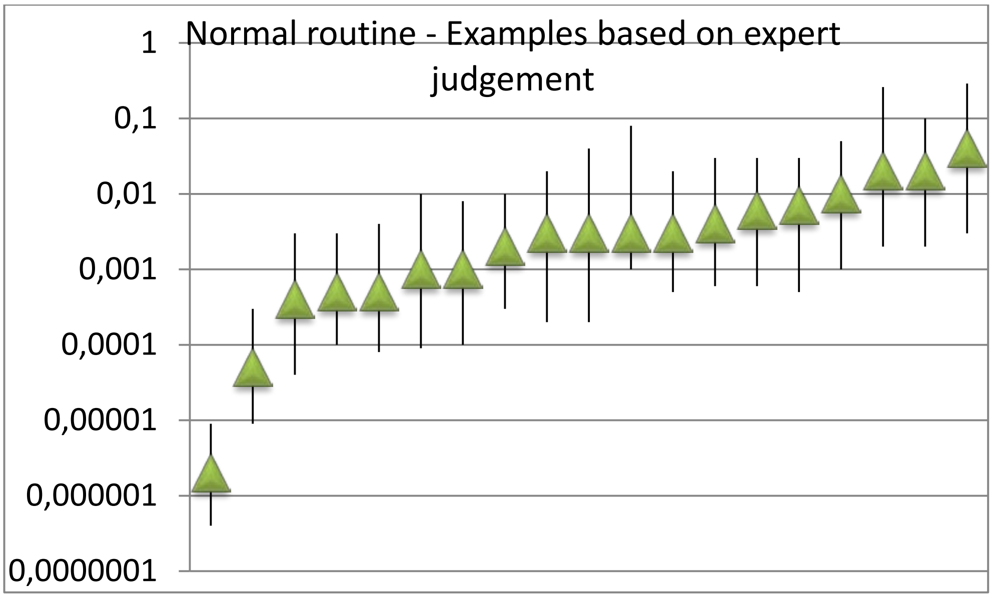
From this set, two different sets are chosen to provide examples of familiar and frequent tasks:

1) **Expert judgments.** A study based on expert judgments by 19 experts with an average of eleven years of experience (Comer et al. 1984) was selected: 19 records, all indicated as highly proceduralised (value=2, see p. 7).

2) **Weekly or more frequently.** Tasks that are performed weekly or more frequent, highly proceduralised (1 or 2) and not based on expert judgments. For this set, those with task frequency (see p.7) 'hourly', 'daily' or 'weekly' were selected. Furthermore, six records were based on expert judgment and these were excluded to distinguish clearly from the previous set. This selection results in a total of <u>22 records</u>.

1. Diagram for study based on expert judgments

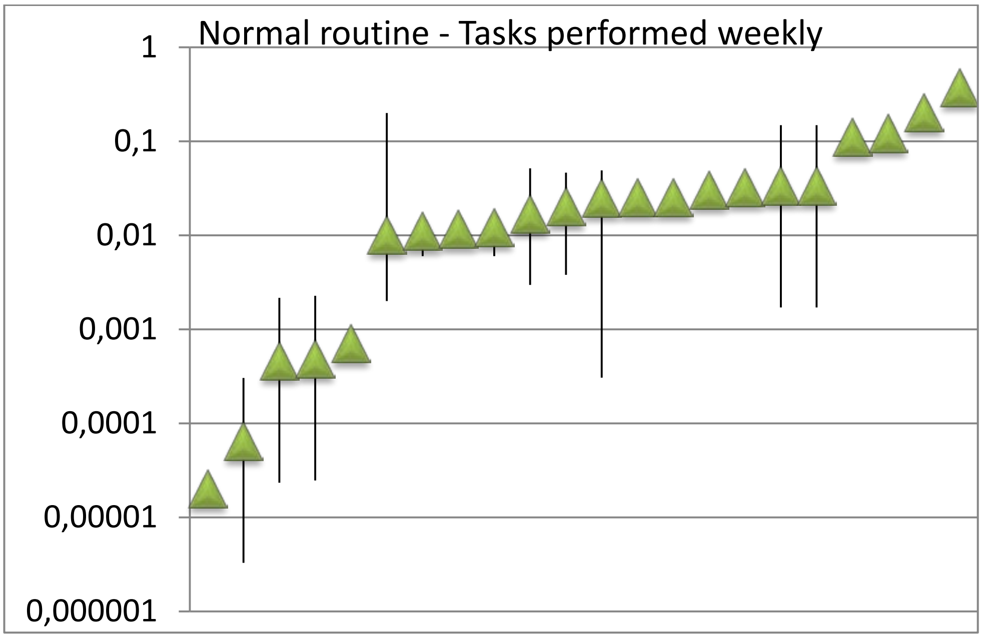
See Appendix B for the list of sampled records (p.35)



Median of HEP: 0.003

2. Diagram for tasks performed weekly or more frequently

See Appendix B for the list of sampled records (p.35).



Median of HEP: 0.022

Comments on estimate, data and method

Medians for both samples are higher than the suggested estimate, the first sample by a factor of almost 8, the second sample ('weekly') by a factor of 50.

The first sample is limited to a specific study of a narrow range of tasks with no consideration for opportunities for recovery. The selection for the second sample is also limited, because only few records in COREDATA have direct information about frequency (parameter: 'familiarity'). The sample may thus be biased, with undue focus on a minority of studies where this parameter is used, e.g. many examples of input failure when using a keypad. There is also a general difficulty in comparing *error* types with *task* types (see previous discussion).

A more appropriate sample would require a more detailed analysis of records (titles and summaries), as well as a careful exclusion of outliers, e.g. 'not recover from error', which is related to a previous error and thus part of a cutest. However, such analysis is beyond the scope of this report.

In any case, the provided estimate is reasonably based on the references and justified by the first sample.

Generic task 2: Nontrivial, familiar

Definition

Familiar and relatively frequent task performed at the knowledge-based level; hence requires judgment.

Comparable to what Banedanmark has described as: normal tasks performed fairly often, but nevertheless a deviation from planned operations.

The task is by default not stressful or urgent; but time pressure may be added as a PSF. It is should be noted that time pressure alone may not necessarily involve stress.

Voice communication tasks are not included.

Emphasis on deviation and judgment indicates a less firm structure with less clearly applicable rules than task 1, which implies different conditions for application of the corresponding PSF: Job aid. This does not mean that there are *fewer* rules for this task type, since degraded operation tends to elicit a lot of rules (at least as far as anticipated situations are concerned).

Higher complexity than task 1.

References

Two GTTs in HEART (see table A.1) indicate various types of degradation from routine operation:

- D (simple, rapid or scant attention; HEP=9E-2)
- C (complex, high comprehension/skill; HEP=0.16).

In the definition of task D, the term 'rapid' indicates time pressure (defined as a PSF in this report), while 'simple' on the other hand indicates a task less complicated than task2. In other words: **D ≅ task2*simple*TP**, where 'simple' is < 1 (improving) and TP > 1 (deteriorating).

Hickling suggests a modification of task C as a baseline *cognitive* task (Hickling 2007:63). As already mentioned, this modification (0.01) is based on the assumption that normal tasks include: procedures, supervision/checking, reversibility (see previous discussion), whereas we only presume reversibility.

Yet another example of a slightly degraded operation can be found in Kirwan's generic guideline data (table A.3): '*11. Error in a routine operation where care is required*' (HEP=1E-2).

The mention of 'care' may represent some judgment corresponding to task2. On the other hand, task2 is not 'routine' in the sense of 'automatic' action: **Kirwan11 ≅ task2*routine**, where routine < 1.

Estimates

We shall again prefer HEART estimates for the same reasons as for task 1, following Hicklings suggestion of a modified task C, although with our more modest modifier of 0.1 representing reversibility.

HEP (task2) = 1.6E-2, interval: 0.012-0.028.

Rail examples from Banedanmark

- Signalman plan/establish temporary speed restriction
- Signalman authorising operational trailing
- Signalman manually setting up route protection prior to routing trains on Signalman protected routes
- Signalman applying wind restrictions
- Driver manually controlling brakes under low adhesion conditions
- Driver controlling train in RM (Restricted Mode) and brakes towards End of Authority
- Driver controlling train under on-sight conditions
- Driver isolating onboard equipment
- PICOP request impromptu possession on handheld terminal
- Signalman plan/apply impromptu possession
- Driver apply clamp to point
- Signalman determine conditions for resetting axel counter
- Signalman applying system override
- O&M Coordinator de-activating low adhesion area

Sampled records

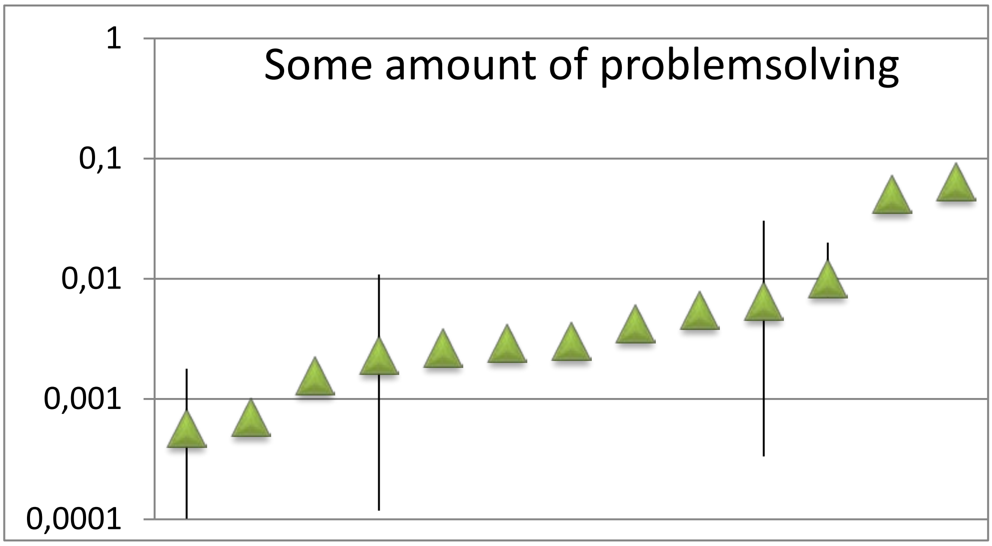
Selection criteria

Records are selected from the non-emergency, non-communication set of 205 records (presented above).

As examples of nontrivial tasks, tasks with some degree of problem solving have been selected. Out of 205 records (without communication studies or emergency tasks), none were rated 5, two were rated 4, and 11 were rated 3. This results in a total of 13 records.

See Appendix B for the list of sampled records (p.36).

Diagram showing distribution of modes and probability intervals



Median of HEP: 0.003

Comments on estimate, data and method

HEP is lower (5 times) than expected – equal to ‘normal routine’ samples based on expert judgment. A reason may be that the selection based on ‘problem solving’ parameter is limited, since only few records have this information.

The sample could be improved (not within the scope of this report) by analyzing records by titles and task summaries to select more relevant examples, or making a search based on ‘cognitive error modes’ (see p.7), selecting errors in ‘decision making’. However, a quick search revealed, against our expectations, that the two dimensions – ‘problem solving’ and ‘decision making’ – are not directly compatible in the database: tasks indicated with low degree of problem solving are also indicated as failures in decision making. This search indicates that we cannot rely solely on either of the parameters, but will have to analyze titles and tasks summaries in any case.

In any case, our estimate is reasonably based on the references, and results from the sample are not dramatically lower.

Generic task 3: Communication, routine

Definition

Communication about familiar contents frequently conveyed that is carried by voice and typically two-way. Some limited template for communication may be available, e.g. predefined ‘elements’ such as identification, message and readback. Error capture and correction is possible (hence, error estimates refer to uncorrected errors).

Note: Communication tasks (both 3 and 4) may often overlap with other generic tasks, both the two emergency tasks 5 and 6 and non-emergency tasks 1 and 2. In general, it is recommended to isolate communication from other elements of a larger task. For example, communication and emergency tasks intersect, and emergency communication might therefore be classified either within communication or emergency generic tasks. We recommend selecting communication as generic task, and then add PSFs (e.g. ‘task competition’ and ‘time pressure’) to account for urgency; however, if communication plays a minor role compared to other task aspects, it will be more appropriate to place such a task within the emergency generic tasks.

This task includes one-way communication in which a message must be transferred from a sender (e.g. signalman) to a receiver (e.g. driver).

Note on PSF job aid: This generic task involves in its definition a procedure or template (phraseology, sequence of communication elements such as caller ID, location etc.), compared to task 4. However, if job aid (physical checklist, written orders) is available, the PSF job aid should be applied.

Rail communication is comparable to Air Traffic Control communication, which is, however, more structured and involves mandatory phraseology to a much greater extent.

References

Communication is often treated as a performance factor rather than a task type, e.g. in the RSSB tool for Human Error *Identification* (Cullen et al 2004:Appendix 5). However, communication ‘factors’ have disappeared in RSSB’s *Quantification* tool, where they are translated to individual factors (Cullen et al 2004:Appendix 11). Yet Cullen et al also have a performance factor for ‘Information quality & availability’ – “*of information given in procedure or via verbal communication (or other method), including shift handover*” = *3 (Cullen et al 2004:122).

However, an appropriate estimate for communication tasks can be found in the CARA (Controller Action Reliability Assessment) tool, which is an application of HEART for air traffic safety management (Kirwan & Gibson 2007):

'9 *communication of safety critical data*', HEP = 6 E-3 – with estimate based on data from EUROCONTROL real time simulation (Kirwan & Gibson 2007:209).

Note that air traffic communication is highly structured, routine and basically one-way, whereas railway communication only occurs in exceptional situations, traffic normally being automatically regulated by signalling systems. Nevertheless, identical estimates has also been applied for other areas, e.g. in NARA (Nuclear Action Reliability Assessment) (Kirwan et al 2004;Kirwan 2008:14).

Estimates

We suggest the estimate from NARA and CARA from the above sources:

HEP (task3) = 6E-3.

It may be considered modifying this estimate, increasing it to take account of the lower degree of structure and routine in railway communication, in particular the low degree implied in task 3. Yet it is difficult to justify a specific modifier.

Rail examples from Banedanmark

- Signalman receive information on need for speed restriction
- Signalman/PICOP communication on planned possession
- Driver reporting location
- Driver receiving timetable update
- PICOP report changes to driving conditions
- Signalman routes call to appropriate Signalman
- Identify receiver/sender using train radio
- Signalman to signalman handover
- Signalman authorising Driver to reverse within the authorised reverse limit (platform overrun)

Sampled records

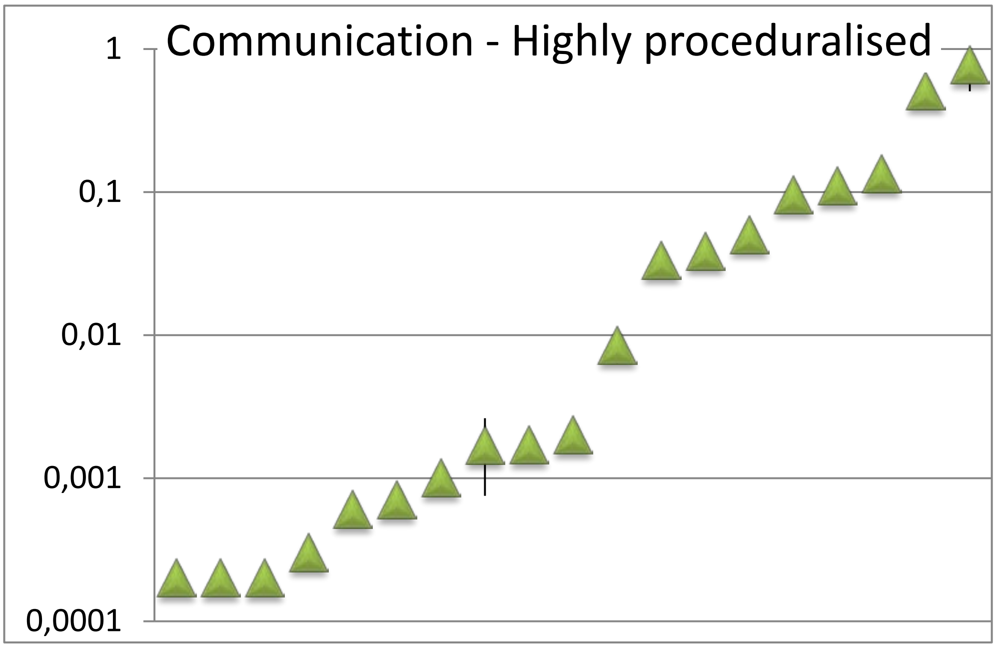
Selection criteria

Communication: Examples of this task and generic task 4 are communication failures identified as records from studies of communication (see p.7). In total, 116 records.

Routine communication records are identified by selecting tasks that are rated as highly proceduralised ('1' or '2', see previous discussion), yielding 19 records.

See Appendix B for the list of sampled records (p. 36).

Diagram showing distribution of modes and probability intervals



Median of HEP=0.002

Comments on estimate, data and method

The median for the sample is lower (3 times) – better – than the estimate provided. This may be because most of the errors are related to highly structured, routine communication tasks.[2]

Yet, various characteristics suggest that the sample does not match the described task ideally. In these communication studies, errors are often identified at a very detailed level, including errors that are often corrected later in a transaction – and thus not relevant with focus on unrecovered errors. On the other hand, the 'proceduralisation' parameter (see previously) is not indicated for most of the records, which may have excluded relevant examples from our sample.

Further selection (not possible within the scope of this report) for this task should therefore be based on analysis of titles and task summaries to allow a focus on unrecovered errors, but also include further examples (without indication of 'proceduralisation'). A quick search shows that studies of Air Traffic Control (included in COREDATA) report a frequency of 1E-3 for 'unrecovered error in pilot readback' (Cardosi 1994), where 'Readback errors' can be taken as an indication of a failure in the original message.

While the sample may thus be improved, the results are nevertheless reasonably compatible with the estimate provided.

Generic task 4: Communication, nonroutine

Definition

Unforeseen, novel content, where no template – or only a rudimentary one – is available.

Includes more interactive types of communication and is thus less 'one-way' than the transmission (reporting/receiving) of a message.

[2] Many records (in COREDATA) with a relatively high HEP – higher than the estimate suggested by Kirwan – come from the same study (Morrow et al. 1993).

Less structured than task 3 and thus a lower knowledge of what to communicate, what to expect the other party to state – less knowledge of the context of the other party and therefore less ability to understand what must be communicated to promote understanding.

May include communication about emergency situation, which would imply some level of stress, compared with similar, non-emergency tasks (see discussion under task 3)

References

Kirwan (table A.3): '*7. General error rate for oral communication*', HEP = 3E-2.

Estimates

Our estimate is directly based on Kirwan's example, which is higher than for task 3, as expected.

HEP (task 4) = 3E-2.

Rail examples from Banedanmark

- Driver reporting unauthorised evacuation of train
- Driver reporting impact with unknown object
- Non-railway competent person calling -Signalman (e.g. emergency services)
- Signalman receiving request for impromptu possession
- Railway emergency call

Sampled records

Selection criteria

Out of 116 communication records, tasks with some degree of problem solving ('4' or '5') are selected: yielding 3 records.

See Appendix B for the list of sampled records (p. 37).

Diagram showing distribution of modes and probability intervals

Neither diagram nor mean value is relevant for only three values.

Comments on estimate, data and method

We have not been able to select an adequate sample to provide examples for task 4. This is because data in COREDATA are from studies of highly structured communication, with no or very few examples of unstructured, nonroutine communication. Relevant examples might be found in other types of studies, although it may be expected that the definition of 'error' is more controversial for less structured communication. Further studies of railway communication are recommended to provide relevant data.

The estimate is thus neither justified nor challenged by the data, but remains a reasonable fit with our task description.

Generic task 5: Emergency scenarios – known

General note on emergency tasks 5 and 6

Generic tasks 5 and 6 are both emergency tasks. They share the essential task aspects of urgency and deviation from normal operating conditions, but they are distinguished by two dimensions: 1) level of urgency (which is related to 'time pressure', with the additional aspect that negative outcome will ensue if not performed correctly on time) and 2) availability of a script and relevant information. Urgency is related to either safety or production pressure, and the latter may in turn lead to safety risks.

Definition

Fairly simple task characterized by some urgency and stress due to safety or production concerns, for which a plan of action (a script) and relevant information are available. May also include 'complicated non-routine' and 'unfamiliar/undefined' tasks that are mission critical (task failure may involve major safety or efficiency breaches).

These characteristics overlaps somewhat with two PSFs, *time pressure* and *job aid*. Stress is not identical to time pressure, but implies that there already is some time pressure in the task. The PSF should therefore only be used with caution, and never 100%. Preferably, increased time pressure may give reason to consider generic task 6 instead. Similarly, availability of a script implies that some 'job aid' (a PSF) is already included in task architecture, which and task 5 is therefore not susceptible to improvement corresponding to a 100 % multiplier.

References

Two relevant examples from Kirwan's 'generic guideline data' (table A.3):

- Kirwan 2: "*Complicated non-routine task, with stress*", HEP=3E-1.
- Kirwan 5: "*Operator fails to act correctly in the first 30 minutes of a stressful emergency situation*", HEP = 1E-1.

While the latter, 'Kirwan 5', does emphasize emergency, the specification of a time frame renders a match with our task definitions quite difficult. The first, 'Kirwan 2', emphasizes stress but seems to imply more complexity than indicated for task 5.

Estimates

We suggest a modified version of the 'Kirwan 2' task mentioned above. We modify to take account of lower degree of complexity, and because estimate for task 5 must be better than task 6:

<u>1E-1</u>

Rail examples from Banedanmark

- Signalman reaction to report of a person being struck by a train
- Signalman apply emergency stop to specified area
- Driver reacting on instruction to bring train to a standstill immediately
- Authorised evacuation of train
- Driver observing unauthorised person on the tracks
- Driver reaction to impact with an object
- Signalman identifying the need for an emergency catenary isolation
- Signalman response to broken overhead wire
- Signalman applying emergency stop to individual train

Sampled records

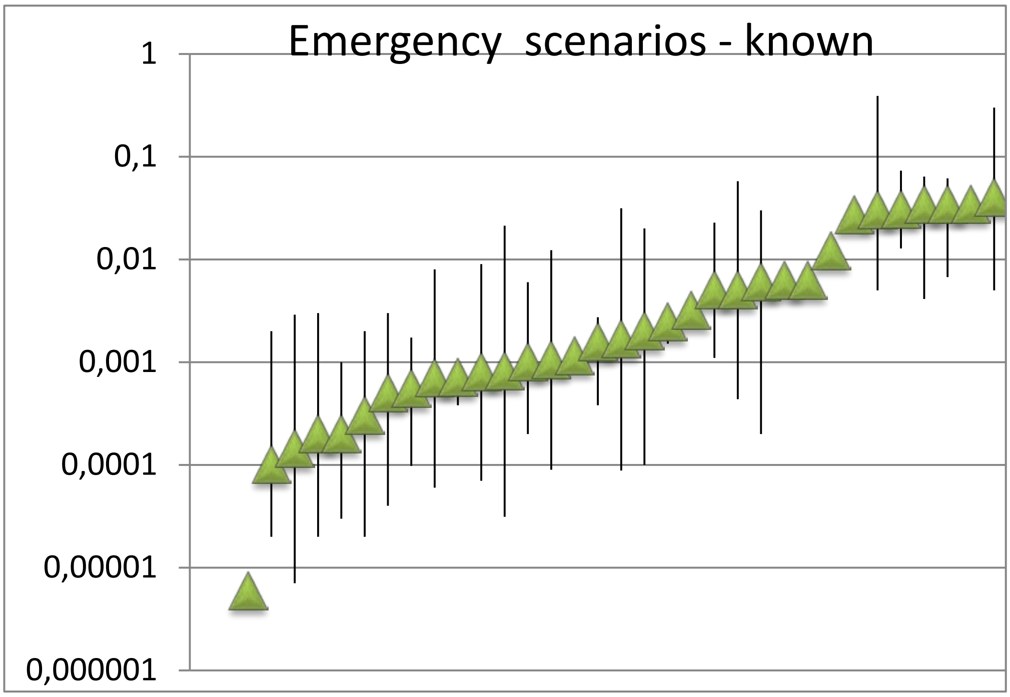
Selection criteria

Selection of examples of this task and generic task 6 (unknown) are based on 93 records with emergency tasks (see p.7).

Highly proceduralised (proceduralisation ≤ 2, see previously) tasks are selected as examples of 'Emergency-known', yielding a total of: <u>35 records</u>.

See Appendix B for the list of sampled records (p. 37).

Diagram showing distribution of modes and probability intervals



Median of HEP: 0.0235

Comments on estimate, data and method

The median for this sample is lower than the suggested estimate for this task by a factor 4. This may be because many failures in the sample are for very simple subtasks in an emergency scenario. Furthermore, most examples are from the same study of a very specific scenario, "Offshore Lifeboat evacuation", making the sample very narrow and less representative. Nevertheless, it does indicate that the estimate arrived at is fairly high, which also confirms the concern that error rate for task 5 should be sufficiently low to 'leave room' for task 6 as the most error-prone one. However, for lack of further references for this task we are obliged to maintain the estimate suggested above. Hickling also provides an estimate of a generic task which is somewhat overlapping with our generic task 5, namely his "*complex task requiring a high level of understanding and skill*" (Hickling 2007: Table 3, p.65).

Generic task 6: Emergency scenarios – unknown

Definition

Task characterized by high degree of urgency and stress due to safety or production concerns, where no adequate plan of action (a script) is available, and relevant information is uncertain or missing.

It differs from task 5 both by the absence of script and scarcity of relevant information, and by increased urgency – related to either safety or production pressure.

The definition of this task already includes error-producing PSFs. On the other hand, there is moderate potential for improvement by adding supervision. (However, compare remarks in the introduction about establishing resilient organizations).

References

Two estimates from the literature seem relevant, the first of Kirwan's 'generic guideline data' (table A.3): "*1. General rate for errors involving very high stress levels*" (HEP=3E-1) and GTT A in HEART (see table A.1): "*Totally unfamiliar, performed at speed with no real idea of likely consequences*" (HEP=0.55).

Estimates

We suggest the lower of the two values as the estimate:

3E-1

Rail examples from Banedanmark

- Signalman reaction to report of unspecified disaster
- Unauthorised evacuation of train initiated by passengers
- Major system breakdown – Signalmen lost indications on signalling control display
- Signalman to identify need to stop traffic in neighbouring infrastructure

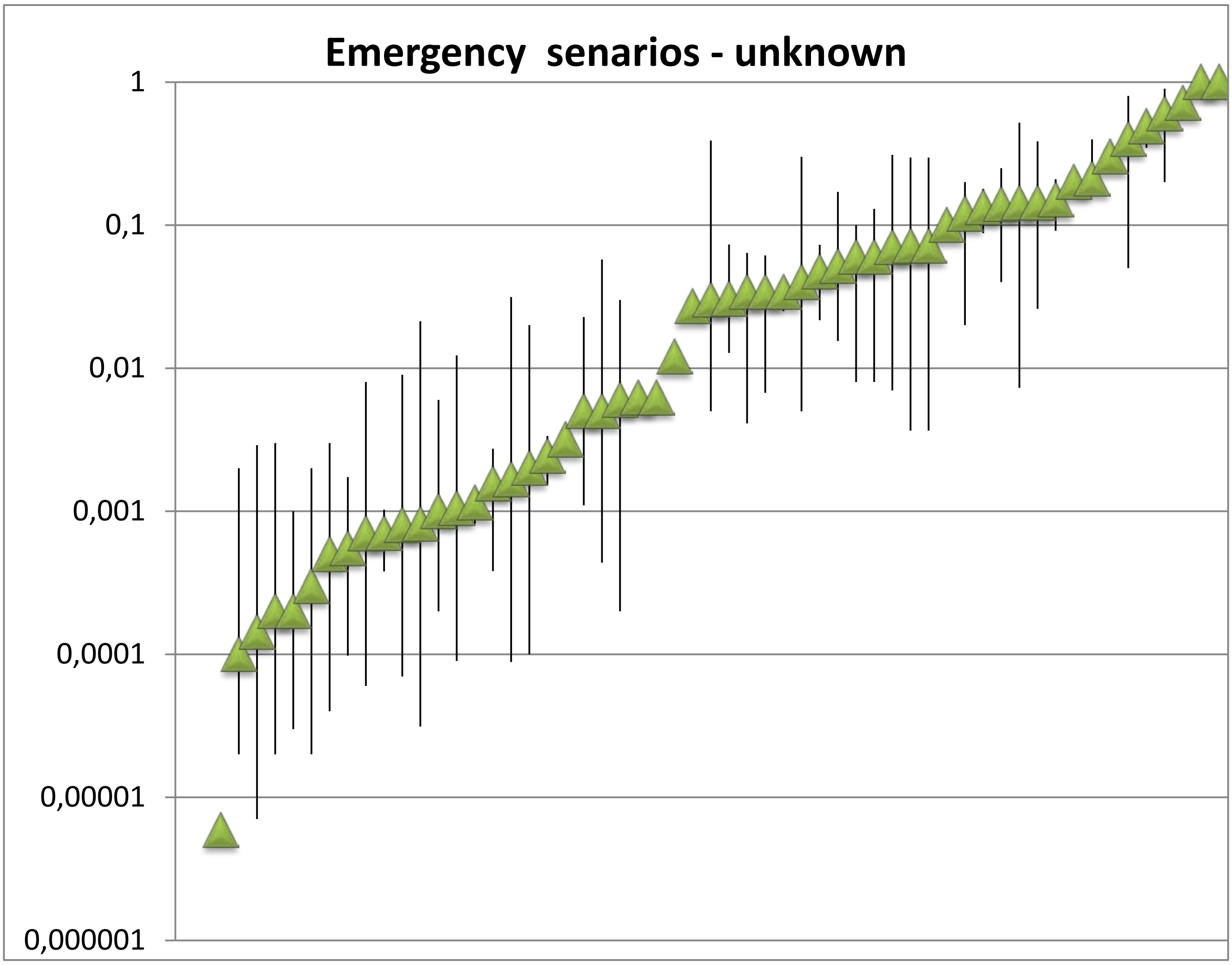
Sampled records

Selection criteria

Based on the same set of 93 emergency records (see previous task), tasks indicated with some degree of problem solving (3 to 5) are selected: a total of 58 records.

See Appendix B for the list of sampled records (p. 38).

Diagram showing distribution of modes and probability intervals



Median of HEP: 0.02851

Comments on estimate, data and method

The median for these examples is lower than expected for ‘emergency - unknown’ tasks – not much higher than the previous estimate for ‘emergency – known’. This may be because many examples are quite well-defined tasks. Only very few are indicated with the highest degree of problem solving, which suggests that the use of this parameter is insufficient to identify tasks without script or relevant information. Furthermore, several examples are parts of cutsets, representing errors related to and/or depending on other errors, and thus with a complicated relation to specific tasks.

It may be possible to make a further analysis and selection based on titles and task summaries, and comparison with examples provided by Banedanmark, possibly by including other studies, but this would be outside the scope of the present report.

We maintain the estimate suggested, since it is based on references, and the definition of task 6 seems to imply a higher error rate than most of the tasks from the sample.

2. Performance shaping factors

There is an overlap between Generic Task Types and Performance Shaping Factors (PSF)[3]. Some factors can be defined either as a parameter in the definition of task types (and the distinction between tasks) or as PSF, for instance “Task competition”. The allocation of such factors either to task types or to a PSF is a matter of choice. But one should be careful – in definitions and/or use – to avoid counting the same factor twice (both as part of the task and as PSF).

There are various reasons for using certain parameters (factors) as PSFs rather than task definition.

1) To describe factors that are *external* to the task itself
2) To isolate characteristics common to several task types. This will allow a more simple definition of task types, with fewer types

Factors considered but chosen as task parameters

Familiarity is regarded as depending on the **frequency** of the task – and thus both as task parameters.

Stress is treated as part of (some of) the generic task descriptions. The immediate BDK use of the parameter is for procedure and system design with an assumption on a standard human performance capability. Thus stress can be counted as an inherent factor of the tasks and not considered as a PSF. For specific task analysis in e.g. incident investigation and analysis this reduction would probably not be advisable but for the intended use it can be justified.

Factors considered but not included

- Level of **Skill (requirement)** and **Experience** will affect operator performance. However, the method presented here will be used in the design of a new system, where it is assumed that operators are qualified for the tasks they perform. Challenges with matching tasks to skills and experience may be considered at a later occasion requiring a more detailed analysis.
- **Procedures**: Standard Operating Procedures (SOPs) are a common method of enhancing task performance in all safety-critical industries. They certainly aid the reliability of simple and sequential tasks, but also have their limitations when operators face uncertainties and ill-defined contexts requiring judgment and situation assessment. In our list of PSFs we have not included procedures, because Banedanmark tasks of type 1, 3 and 5 must be assumed to be supported by procedures.

Note on application of PSFs

A PSF is normally applied by multiplying HEP for a specific task with the appropriate multiplier, e.g. an ‘emergency – known’ may be improved by supervision, resulting in a HEP = 0.5 * 1E-1 = 5E-2. Often, however, it will not be reasonable to use the ‘full’ multiplier, either because the PSF is only partly active, e.g. limited time pressure or a simple job aid, or because the PSF is already partly implied in the generic task definition. In this case, a ‘partial’ multiplier can be applied, using a value between 1 (no effect) and the full multiplier (> 1 for negative PSFs, <1 for positive PSFs). For example, a simple routine task may occur under medium time pressure: HEP = 2 E-3 * 5 = 1 E-2.

Performance Shaping Factors selected for Banedanmark

Four performance factors have been selected in agreement with Banedanmark. Two of the PSFs – time pressure and task competition – represent negative conditions that can increase error rates, while the other

[3] Performance Shaping Factors are also referred to by other labels in the HRA literature, such as Performance Factors (PF) (Cullen et al 2004) or Error Producing Conditions (EPC) in HEART (Kirwan 1994;Williams 1986).

two – supervision and job aid – represent potential improvements that can be added to reduce error rate but will require an investment of resources (e.g. extra use of personnel or extra equipment).

Table of Performance Shaping Factors (PSFs)

Performance Shaping Factors				
Titel	**Explanation**	**Multi-plier**	**References**	**Caution for 'double counting'** *Tasks, PSFs*
1. Supervision	Having a supervisor who monitors or who is in the background or on call will generally improve reliability and thus reduce the error rate.	**0.5**	(Dickinson & Lowe 2007;Weber 1992)	
2. Job aids	Different types of aids or tools will support the operator and reduce the error rate. E.g. a checklist to support memory; alarm to direct attention; a fixed format for communication may reduce omissions or misunderstandings	**0.1**	(Cullen et al 2004)	3;5
3. Task com-petition	Error rates increase when an operator must perform more than one task at the same time, e.g. when being distracted by other tasks and thus forced to prioritize	**6**	(Cullen et al 2004)	PSF 4
4. Time pres-sure	Error rates increase when operator is under time pressure	**11**	(Williams 1986)	5;6; PSF 3

PSF 1 - Supervision

Definition

Having a supervisor who monitors or who is in the background or on call will generally improve reliability and thus reduce the error rate. The supervisor can check and correct actions, but also has other roles, e.g. by providing 'focused attention/supervision' (Dickinson & Lowe 2007:227) on critical events, thus relieving operators in emergency situations (tasks 5 and 6) or reduce the effect of 'task competition' (PSF).

Discussion

There is a potential overlap/conflict with PSF 'time pressure' in the sense that some benefits of supervision depend on adequate time to check and intervene. High 'time pressure' will reduce this benefit of supervision.

References

Supervision can be identified in HEART as one of the factors implied in the 'multiplier' identified by comparing GTTs B and F (see previously). The 'B/F multiplier', ca. 1/90. thus represents a product of three factors: *supervision*, *procedures* and *reversibility*.

According to Weber, supervision reduces errors by 50%: "*it has been suggested that 50 % of unsafe acts could be eliminated by the use of effective supervisors*" (Weber 1992) referred in (Dickinson & Lowe 2007:225).

HEART EPC (negative): "*17. Little or no independent checking or testing of output*": *3 (table A.2).

Estimated value

We recommend the value suggested by Weber: 0.5.

It seems fair to assume that the two other elements in the 'B/F multiplier', *reversibility* and *procedures*, are somewhat more effective – in order to arrive at a total product of 1/90.

Rail examples from Banedanmark

- Second person available to correct actions in real time, e.g. second Driver repeating restrictive indications on the DMI
- Verification by other person, e.g. Signalman verifying O&M coordinator planning of temporary speed restriction

PSF 2 - Job aids

Definition

A job aid is an external, physical or electronic artifact which is associated with a procedure such that the artifact and the procedure support the execution of a task. All job aids are associated with one or several procedures, but not all procedures are accompanied by a job aid. A job aid can be in the form of a checklist that supports memory; an alarm that aids supervision tasks or brings attention to a task that requires immediate attention, etc.

Discussion

While job aids typically have a positive impact on performance and reliability, we may also take into account sources treating them as a negative PSF. Procedures can influence performance negatively in various: lacking where demanded, unclear, outdated, frequent changes, contradicting (Cullen et al 2004:132;Kirwan 1994:228). In this case, we focus on the difference between procedures and "*no procedures, even though the task needs them" (Kirwan 1994:228)* or procedures that *"do not specify clearly what the driver is required to do (e.g. no procedure in place for a particular, unusual circumstance)"* (Cullen et al 2004:132), while assuming that they are not unclear or conflicting, since the method presented in this report is to be used for the design of a future system. Nevertheless, one should be aware that not all tasks can simply be improved by a job aid that implies a specific structuring, e.g. if the implied structure is too inflexible to fit the context.

The ambiguous benefits of job aids are also indicated by the fact that 'inadequate human machine interfaces' are also treated as a negative PSF (Kirwan 2008:12). For the purposes of this report, we shall assume that a planned job aid will be adequate, though with a caution against regarding it as a universal medicine, and a preference for a modest estimate.

References

'Job aid' seems to constitute the main difference between two GTTs in HEART (table A.1):

- G: "*without the benefit of significant job aids*", HEP=4E-4),
- H: "*an augmented or automated supervisory system providing accurate interpretation of system stage*", HEP=2E-5.

The main difference between the two task is "*automated support*" (Hickling 2007:63) or a very significant job aid, representing an improvement multiplier of = 20.

A considerably less powerful multiplier may be identified from a negative PSF in the RSSB tool: ‘No progress tracking’, explained as “*no means (e.g. checklist) of tracking progress on a sequence of tasks*”, increasing error rates by 1.4 (Cullen et al 2004:124). This is a *negative* multiplier representing lack of ‘means’, whereas the availability of such means (a particular type of job aid) would improve error rates by the inverse factor.

The RSSB tool has PSF multiplier for (lack of) ‘Procedures clarity’ = 5 (Cullen et al 2004:122). While the full multiplier represents conflicting procedures, a reduced value may probably be applied for ‘missing procedures’. Problems with procedures are also mentioned in another PF in the RSSB tool: ‘Information quality & availability’ – “*of information given in procedure …*”: 3 (Cullen et al 2004:122).

Estimated value

We recommend a relatively modest multiplier, at a value that is half the one derived from comparing the two HEART tasks.

Job aid multiplier = <u>0.1</u>.

For any particular case it must be assessed whether full multiplier weight should be applied, since the effect of job aids vary considerably.

Rail examples from Banedanmark

- Written Orders to assist message exchange
- Reminder applications, e.g. indications of active possessions, temporary speed restrictions etc.
- Checklists
- Decision support system
- Dual-faced stop markers indication work site limit
- Handheld terminal to support PICOP requesting track possession

PSF 3 - Task competition

Definition

Task competition applies when an operator must perform more than one task at the same time. This also includes situations when an operator is distracted by other tasks and thus forced to prioritize.

Discussion

Task competition differs from ‘simple’ distraction, which implies a clear difference and priority between primary task and distracters: in ‘task competitions’, there is no clear priority among the competing tasks. It implies a higher workload/time pressure, but with the extra challenge from the simultaneousness and unclear priority that complicates queuing of tasks.

References

Distraction seems implied in a task type in Kirwan (Table A.3): '*4. Non-routine operation, with other duties at the same time*', HEP=1E-1, although he does not provide a comparable estimate for the ‘complementary’ ‘non-routine operation’, which would have allowed calculating a multiplier for ‘other duties at the same time’.

The RSSB tool has a PSF multiplier = 6 for 'High workload', the definition also emphasizing multi-tasking: "*High levels of driver workload due to multi-tasking requirements, disruption/faults, DOO operations, complex route etc.*" = *6 (Table A.4)[4].

The RSSB also has a very modest PSF multiplier for 'Distraction/3rd parties' (RSSB tool) = 1.03 (Cullen et al 2004:124) (Table A.4).This multiplier is so low and insignificant that one could suspect a mistake? In any case, the examples provided generally suggest secondary distracters rather than competing tasks. On the other hand, part of the definition suggests otherwise: "*preoccupied by an issue of concern*" (although that issue might be private rather than work-related).

Estimated value

We recommend using the RSSB multiplier for 'high workload', although that definition implies a caution against using it together with PSF time pressure.

Task competition multiplier = <u>6</u>

Rail examples from Banedanmark

- Several duties at once e.g. dispatching and traffic information
- More than one railway emergency

PSF 4 - Time pressure

Definition

Time pressure is an active performance shaping factor whenever an operator must perform one or several tasks at a rate which is close to the operator's time/error performance ceiling.

Discussion

Time pressure corresponds to a higher workload, but does not imply simultaneous tasks with unclear priorities (see PSF 3).

Time pressure means less time for error detection and correction. On the one hand, this reduces some benefits of PSF 'supervision'. On the other hand, such a supervisor could relieve the workload by performing some of the tasks.

It may be assumed that some time pressure already exists in emergency tasks – therefore, this PSF cannot be applied fully to tasks 5 and 6.

References

Time as parameter ('speed', 'rapidly', 'time to correct') may be identified in the definitions of most GTTs in HEART (A,D,E,G), but here is an appropriate EPC: "*2. A shortage of time available for error detection and correction*" (Williams 1986) (Table A.2), which was also used directly in the RSSB tool (table A.4).

Estimated value

We recommend using the HEART EPC directly:

PSF time pressure = <u>11.</u>

---

[4] This is probably an adaptation of HEART EPC "*8 A channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information*" (table A.2), although this definition specifically emphasizes information overload.

Rail examples from Banedanmark

- Immediate evacuation required due to unsafe situation on a train
- External conditions not yet in place to allow safe evacuation (e.g. broken overhead wire in the vicinity of a train on fire).
- Build up of tasks due to follow-on deviation resulting from initial deviation from plan.

Appendix A. Supporting tables

Table A.1 HEART Generic categories

Eight generic categories, A-H, in HEART. From (Kirwan 1994, p.237)

Generic task	Proposed nominal human unreliability (5th–95th percentile bounds)
(A) Totally unfamiliar, performed at speed with no real idea of likely consequences	0.55 (0.35–0.97)

[redacted]

Table A.2 HEART Error Producing Conditions (EPCs)

Twenty-six error-producing conditions defined in HEART. EPCs are similar to Performance Shaping Factors. From (Kirwan 1994, p.238).

Error-producing condition	Maximum predicted nominal amount by which unreliability might change, going from 'good' conditions to 'bad'
1. Unfamiliarity with a situation which is potentially important [redacted] but which only occurs infrequently, or which is novel	× 17
2. A shortage of time available for error detection and correction	× 11
3. A low signal-to-noise ratio	×10
4. A means of suppressing or overriding information or features which is too easily accessible	×9
5. No means of conveying spatial and functional information to operators in a form which they can readily assimilate	×8
6. A mismatch between an operator's model of the world and that imagined by a designer	×8
7. No obvious means of reversing an unintended action	×8
8. A channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information	×6

Table A.3 Kirwan generic guideline data.

A table of eighteen generic guideline data presented by Kirwan. From (Kirwan 1994, p.379)

Description	Human-error probability
1. General rate for errors involving very high stress levels	0.3
2. Complicated non-routine task, with stress	0.3
3. Supervisor does not recognise the operator's error	0.1

[redacted]

Table A.4 Performance Factors from RSSB Human Error Quantification

(Cullen et al 2004:Appendix 10)

Ref	Performance factor	Multiplier	Definition	Examples
1	Unfamiliarity	17	Driver is COMPLETELY unfamiliar with routes, cabs, procedures, scenarios, jargon (if not completely unfamiliar use 12 – Driver experience). This is used mainly for emergencies and major disruption and applies to rolling stock, routes, procedures etc.	• Driver used to driving on mainline may not have encountered ground position light signals in sidings • Driver not practiced in using pass-com system effectively. • Driver fails to correctly follow unfamiliar procedure in emergency situation
2	Time availability	11	Lack of time available to the driver to complete a job or task e.g. as a result of delays, disruption or engineering works.	• Signal failures cause delays to journey meaning that driver may have to lodge for the night instead of returning home. • Insufficient time scheduled into diagrams to properly carry out all required checks before taking a train. • Express train routinely delayed by Signallers giving priority to freight trains and stopping services • Insufficient time available for driver to lay protection on track in an emergency
3	Ability to detect and perceive	10	Inability to adequately perceive visual and auditory signals (signals, speed boards, alarms, radio). Applies to cases of poor visibility due to poor signal sighting or weather conditions, or where auditory signals are difficult to detect due to background noise.	• Flood light outside signal box causes glare for driver rendering signal difficult to see. • Noisy cab environment makes audible warnings more difficult to hear. • Dirty speed restriction warning boards. • Weather conditions e.g. fog, bright sunlight, heavy rain, snow • Overgrown vegetation shortens signal sighting distance • Station features, e.g. clocks, hanging baskets in front of off indicators
4	Features over-ride allowed	9	No interlocking/ time out facility to prevent features being over-ridden i.e. too easy for driver to inappropriately suppress or circumvent information or input from safety features.	• TPWS can be temporarily over-ridden e.g. when passing red signal on Signaller's authority or when passing red signals in a possession
5	Positioning and layout	5	Positioning of equipment does not reflect how it will be used (e.g. according to frequency and order of use, importance of piece of equipment, functionally grouped etc). Does not apply to cases of single signal sighting (in these cases use 3 – Ability to detect and perceive).	• Lever is pushed forward to take power in some units and pulled back in others. • Door opening buttons for left and right sides of train are not logically arranged. • Driver inputs wrong code into radio by mistyping due to poor layout of keypad
6	High workload	6	High levels of driver workload due to multi-tasking requirements, disruption/ faults, DOO operations, complex route etc.	• Driver negotiating complex route fails to notice speed restriction warning board. • Driver dealing with passenger incident via passcom fails to utilise Driver Reminder Appliance at station/ red signal. • Driver has many tasks to perform at stations. • Rush hour if DOO (otherwise driver probably unaffected) • Too much paperwork to complete

Ref	Performance factor	Multiplier	Definition	Examples
7	Technique learning	6	Driver is required to apply a new technique which requires an opposing philosophy to that with which the driver is familiar.	• Driver opens doors on wrong side of train because all other stops on route have the platform on the opposite side. • Driver fails to stop at a request station that is used infrequently.
8	Procedures clarity	5	Driver is unclear as to what is required of him /her. Applies in cases where there are numerous different procedures and it is not clear which ones apply or when procedures keep changing and driver is unsure whether his/ her knowledge is current.	• Company procedure seems to contradict Rule Book • Procedure is open to misinterpretation • Procedure is missing
9	Risk perception	4	Driver underestimates risk (complacency), possibly due to over-familiarity, or overestimates it (over cautious)	• Driver makes a call on a mobile phone believing it will not compromise the driving task • Driver passes signal at danger at a controlled approach signal because it has always cleared in the past
10	Poor feedback	4	Feedback provided to driver is inadequate/ poorly presented.	• In cab devices do not provide good quality feedback
11	Delayed/ Incomplete feedback	4	Feedback is not provided to driver at required point or is incomplete.	• Driver waiting at red light repeatedly calls Signaller but call is not answered
12	Driver experience	8	Inexperience of driver or unfamiliarity with routes, cabs, procedures, scenarios, jargon. Use 1- Unfamiliarity if driver is completely inexperienced.	• Inappropriate braking due to lack of experience of braking characteristics on new type of rolling stock • Driver is travelling route that he/ she seldom takes • Jargon • Route driver has only recently signed for
13	Information quality & availability	3	Poor quality & availability of information given in procedure or via verbal communication (or other method), including shift handover.	• Driver is unsure of which course of action to take due to Signaller not following protocol during an exchange. • Driver fails to inform Replacement driver about a fault on the train • Problem in understanding regional accent • Poorly presented information such as procedures
14	Objectives conflict	3	Conflict between immediate and long term objectives. E.g. long term objective to get to destination, short term requirement to stick to speed limit; also driver receives mixed message	• Driver fails to check doors carefully when trying to set off from station on time • Driver fails to take faulty train out of service due to pressure to meet punctuality targets • End of speed restriction on approach to red signal
15	No diversity	2.5	Operator has to rely on single source of information for confirmation of accuracy e.g. single source alerts	• Driver knows a TSR is in place from the WON but the Advanced Warning Board is missing • Red signal is indicated only by the actual light itself, there is no AWS, DRA or other information for the driver to use.

Ref	Performance factor	Multiplier	Definition	Examples
16	Education and training	2	Required training qualifications/ levels of competency not possessed by driver; Poor training; Lack of refresher training; Lack of practical training.	• Drivers are expected to sign for a route which they have only watched a videotape of, not driven. • Drivers have not received safety training • Drivers are overdue for practical training • Drivers are trained by someone not competent themselves • Drivers have not received refresher training
17	Trust in information	1.6	Trust the driver has in equipment or person providing information to him/ her; poor equipment reliability	• Signal is known to revert to red causing technical SPADs so drivers approach it at caution unnecessarily. • Speedometer is unreliable • Driver does not trust person giving information • Driver does not trust equipment with history of faults, now reliable
18	Job clarity	1.6	Poor clarity of who is responsible for what i.e. driver does not know his/ her responsibilities and boundaries.	• Driver uncertain as to whether he can refuse to take train if over-crowding is too severe • Driver unsure as to whether he/ she is supposed to report overgrown vegetation
19	Physical capabilities	1.4	Individual's physical capabilities (e.g. physical strength, reach) are not adequate for the task.	• Driver's eyesight has deteriorated to the point where glasses are necessary. • Due to obesity, driver would not be able to run down the track to lay detonators in the event of an emergency
20	Emotional stress	2	Stress and mental health levels due to mental illness; personal problems; threat of assault, incidents; suicides	• Drivers are under threat of redundancy. • Driver has financial problems. • Driver forgets DRA at station due to Manager demanding ID and proficiency cards unexpectedly • Driver fears for his physical safety after assaults or attacks on cab • Driver has experienced a suicide
21	Health	2	Poor driver health including effects due to medication.	• Driver has badly managed diabetes and may be liable to pass out. • Driver has back problems causing discomfort when sitting for long periods. • Driver feels pressured into working when ill due to Managing For Attendance
22	Morale *Note: Choose 28 – Team Relations if problem is between colleagues.*	2	Motivation, trust, respect, degree to which driver feels valued.	• Drivers are poorly paid in relation to drivers from another comparable company. • Driver is involved in ongoing dispute with Manager • Company has introduced unpopular new initiatives
23	Consistency of displays	1.2	Consistency of displays (both in-cab and lineside) in terms of how information is presented.	• Driver has to switch between cabs displaying speed in kph and mph • Non-standard signage on lineside

Ref	Performance factor	Multiplier	Definition	Examples
24	In-cab environment	8	Physical environment discomfort – temperature, lighting, ventilation, noise.	• Cab heater is stuck full on and driver cannot open the window due to the excessive noise. • Cab fascia does not fit well and cold draughts blow around driver's legs. • Audible warning volume is set too loud and startles drivers.
25	Concentration	3	Boredom, monotony, repetition causing possible lapses in concentration.	• Diagram consists of same commuter route eight times in a row. • Express route with few station stops and driver experiences mainly green signals
26	Fatigue	1.1	Fatigue due to factors such as shift design, rest breaks, staffing levels, home life, hot cabs.	• Driver has small children and has been unable to sleep during the day when on night shifts. • Driver has been unable to sleep due to poor lodging accommodation. • Driver fails to respond to ESR board due to fatigue • Fatigued due to heat
27	Distraction/ 3rd parties	1.03	Other people in/ around the cab which may cause distraction; Driver may be distracted by people/ events or preoccupied with an issue of concern.	• Member of railway staff travels to work in driver's cab. • Drunken passengers create noise and bang on driver's cab door. • Driver distracted by station activities or platform distractions • Driver distracted by trackworkers on lineside • Driver distracted by activities adjacent to line
28	Team relations *Note: Choose 22 – Morale if poor relationship is with Manager*	1	Team harmony/ friction between driver, Guard, Signaller etc.	• Driver has to work with an ex-partner • Driver disagrees with Guard over method of despatch
29	No progress tracking	1.4	Driver has no good means (e.g. checklist) of tracking progress on a sequence of tasks.	• No checklist for tasks during train preparation • No checklist for emergency procedures • Driver loses place on route due to fog or featureless landscape

Table A.5 Examples of Performance Shaping Factors cited in first-generation HRA tools

From (Lee et al. 2011):

Table 1
Mapping of PSFs in the first-generation HRA methods.

THERP	SLIM	INTENT	HEART	PSF categories
Physiological stressors, psychological stressors	Competence	Stress, workload	A channel capacity overload, a need for absolute judgments which are beyond the capabilities or experience of an operator	Stress
Task and equipment characteristics	Role of operations	SRK	–	Action type
Organismic factors	–	Experience	Operator inexperience	Experience
Situational characteristics	–	–	A shortage of time available	Time available
Situational characteristics	–	–	No clear, direct and timely confirmation of an intended action	Places where operator action are taken
Job and task characteristics	Meaningfulness of procedures	Procedures	–	Procedures
Organismic factors	–	Training	A need to unlearn a technique and apply, a mismatch between the educational-achievement level of an individual and the requirements of the task, a mismatch between perceived and real task, a conflict between immediate and long-term objectives, little opportunity to exercise mind and body outside the immediate confines of a job	Training
–	Quality of design	HMI	No means of conveying spatial and functional information, poor, ambiguous or ill-matched system feedback, no diversity of information input for veracity checks, unreliable instrumentation, a means of information or features, a mismatch between an operator's model and designer	HMI
–	Teams	Safety culture, communication	No obvious ways to keep track or progress during an activity	Teamwork

Table A.6 Examples of Performance Shaping Factors cited in second-generation HRA tools

From (Lee et al 2011):

Table 2
Mapping of PSFs in second-generation HRA methods.

CREAM	HRMS	SPAR-H	ATHEANA	IDAC	PSF Categories
Number of simultaneous goals	Task complexity	Complexity, stress/stressors	Workload, time pressure, and stress	Emotional arousal	Stress
Number of simultaneous goals	Task complexity	Complexity	Complexity	–	Action type
Adequacy of training and experience	Training/expertise/ experience/competence	Experience/ training	Applicability and suitability of training/ experience	Memorized information	Experience
Available time	Time	Available time	Time available	Strains and feelings	Time available
Working conditions	–	–	Environment, accessibility and operability of the equipment	Environmental factor, conditioning events	Places where operator action are taken
Availability of procedures/Plans	Procedures	Procedures	Suitability of relevant procedures and administrative controls	Organizational factors	Procedure
Adequacy of training and experience	Training/expertise/ experience/competence	Experience/ training	Applicability and suitability of training/ experience	Memorized information	Training
Adequacy of HSI and operational support	Quality of information/ interface	Ergonomics/ HSI	Ergonomic quality of the HSI, availability and clarity of instrumentation	-	HMI
Adequacy of organization, crew collaboration	Task organization	Work process	Team/crew dynamics	Team related Factors	Teamwork

Appendix B. Reference records (COREDATA database)

Sampled records

Generic task 1 – Sample A: Study based on expert judgments

Short_Title
Nuclear Power plant operation - No action on 5 annunciators alarming
Nuclear Power plant operation - Incorrect reading of digital indicator
Nuclear Power plant operation - Wrong choice of valve (adequate labelling)
Nuclear Power plant operation - Wrong switch choice (arranged with clearly drawn mimic lines)
Nuclear Power plant operation - Switch turned in wrong direction
Nuclear Power plant operation - Reading wrong meter (arranged with clearly drawn mimic lines)
Nuclear Power plant operation - Incorrect interpretation of value shown on a chart recorder when the chart recorder has normal bands indicated on the scale
Nuclear Power plant operation - Wrong Switch Choice (grouped according to function)
Nuclear Power plant operation - Select the wrong circuit breaker
Nuclear Power plant operation - Failure to realize that a valve is not in its proper position
Nuclear Power plant operation - Meter out of range not noticed
Nuclear Power plant operation -Setting 10-position rotary selector switch to wrong position
Nuclear Power plant operation - Wrong switch choice (Identified by labels)
Nuclear Power plant operation - Reading wrong meter (grouped according to their function)
Nuclear Power plant operation - Incorrectly reads graph
Nuclear Power plant operation - Reading wrong meter (identified by labels only)
Nuclear Power plant operation - Wrong choice of valve (inadequate labelling)
Nuclear Power plant operation - Jammed meter not recognised
Nuclear Power plant operation - Failure to act on 1 out of 10 alarming annunciators

Generic task 1 – Sample B: Tasks performed weekly or more frequent

Short_Title
Unrecovered omission errors per character entered
Waste handling error
Accidental override of a safety interlock
Nuclear -Waste Handling
Nuclear - Tank Discharge Valves
Optimal training regime
Manipulation of numbers using a pocket calculator
Unrecovered commission errors per character entered
Look-up values in a table

Manufacturing - inadvertent puncturing of bellows in a motor
Non-optimal training regime
Door safety mechanism deliberately disabled
Keying numerics without a short term memory demand
Keying numerics with a short term memory demand - early digits
Dial an incorrect 10 digit number using a conventional keyphone
Errors per character entered
Manufacturing - inadequate coating of a surface
Manufacturing - leak testing
Dial an incorrect 10 digit number using a capacitative telephone key pad
Dial an incorrect 10 digit number using a membrane telephone key pad
Keying numerics with a short term memory demand - later digit
Do not recover from an error

Generic task 2: Nontrivial, familiar

Short_Title
Omit positioning of a control switch
Nuclear - Chemical discharge
Permit to work - fail to identify obscure hazard during preparation of a confined space entry permit
Chemical industry - select an invalid button at a central control panel
Permit to work - inadequate definition of equipment location
Permit to work - complete a task using methods outside the scope of the permit to work
Chemical industry - input an incorrect numeric setpoint at a central control panel
Chemical industry - control action unsuitable for the equipment state
Permit to work - personal protective equipment inadequately specified
Chemical industry - input incorrect alphanumeric equipment identifier at a central control panel
Numeric ranking exercise - provide incorrect answer
Numeric ranking exercise - misunderstanding a question
Air traffic control - touch the wrong part of a touch screen

Generic task 3: Communication, routine

Short_Title
Strip Marking Error
Relative Speeds are Misjudged
Wrong Aircraft Instructed
Climb/Descent Misjudgement
Headings are Misjudged

Readback Error
Unrecovered readback errors per aircraft controlled
Routine air traffic control communication - pilot responds to message intended for another plane
Callsign Confusion
Unrecovered strip marking errors per aircraft controlled
Routine air traffic control communication - pilot error in read back
Routine air traffic control communication - missing acknowledgement
Routine air traffic control communication - partial readback
Unrecovered readback errors per hour
Routine air traffic control communication - missing readback
Unrecovered strip marking errors per hour
Routine air traffic control communication - missing or partial readback
Routine air traffic control communication - controller fails to correct a pilot error in read back
Routine air traffic control communication - controller fails to identify that the wrong pilot has responded to their message

Generic task 4: Communication, nonroutine

Problem solving

Short_Title
Split Sector not Coordinated Adequately
Radar Conflict Alert is Assumed to be a False Alarm
A Loss of Separation is Occurring but ATCO does Nothing

Decision Making

Short_Title
Headings are Misjudged
Climb/Descent Misjudgement
Relative Speeds are Misjudged
Radar Conflict Alert is Assumed to be a False Alarm
A Loss of Separation is Occurring but ATCO does Nothing

Generic task 5: Emergency scenarios – known

Short_Title
Offshore Lifeboat evacuation / Failure to start lifeboat engine at any point in the evacuation CONTROLLED
Offshore Lifeboat evacuation /Failure to start the engine before the boat is lowered CONTROLLED
Nuclear power plant operation - initiate residual heat removal service water (RHRSW) system
Offshore Lifeboat evacuation / Failure to start lifeboat engine at any point in the evacuation SEVERE

Offshore Lifeboat evacuation /Failure to start the engine before the boat is lowered SEVERE
Offshore Lifeboat evacuation / Check wheel is positioned to clear installation CONTROLLED
Offshore Lifeboat evacuation / Check wheel is positioned to clear installation SEVERE
Offshore Lifeboat evacuation / Ensure hatches and doors secure - omitted CONTROLLED
Nuclear power plant operation - restoring offsite power before attempting to restore power using the diesel generation.
Engagement of forward gear and full throttle
Lifeboat driven without disengagement
Hooks not properly disengaged
Hooks reset too early
Offshore Lifeboat evacuation / Early removal of retaining pin
Offshore Lifeboat evacuation / Ensure crew securely strapped in. CONTROLLED
Offshore Lifeboat evacuation / Break cable not operated long enough CONTROLLED
Offshore Lifeboat evacuation / Attempting to drive the boat without disengaging the boat from the hooks CONTROLLED
Offshore Lifeboat evacuation /Ensure that drop zone is clear CONTROLLED
Offshore Lifeboat evacuation / Omit 6 second purge CONTROLLED
Omission of an emergency event task when procedures are available
Omit 6 second purge
Offshore Lifeboat Evacuation / Omit check of fuel tap slackage
Offshore Lifeboat Evacuation / Initiation of air support system
Unable to disengage boat in one quick movement
Operation of break cable too short
Incorrect disengagement of hooks
Offshore Lifeboat evacuation / Ensure hatches and doors secure - omitted SEVERE
Offshore Lifeboat evacuation / Break cable not operated long enough SEVERE
Offshore Lifeboat evacuation /Ensure that drop zone is clear SEVERE
Offshore Lifeboat evacuation / Monitor compass to ensure leaving platform at 90 degrees. CONTROLLED
Offshore Lifeboat evacuation / Attempting to drive the boat without disengaging the boat from the hooks SEVERE
Offshore Lifeboat evacuation / Ensure crew securely strapped in. SEVERE
Omission of wind speed, direction and sea state check
Offshore Lifeboat evacuation / Omit 6 second purge SEVERE
Offshore Lifeboat evacuation / Monitor compass to ensure leaving platform at 90 degrees. SEVERE

Generic task 6: Emergency scenarios – unknown

Short_Title
Nuclear plant - for a well defined mimic layout: failure to recover from selecting and operating the wrong control during an emergency event
Nuclear plant - fail to recover from error of turning a control in the wrong direction (when the correct control direction is opposite to typical operator expectation) during an emergency event
A Loss of Separation is Occurring but ATCO does Nothing
Nuclear power plant operation - failure to initiate the standby liquid control (SLC) during an emergency.

Drilling for oil/gas offshore - pump liquid too slowly through the drilling system leading to a further emergency event
Nuclear power plant operation - Actuate the suppression pool cooling mode during an emergency
Nuclear power plant operation - Scramming the reactor during an emergency.
Nuclear power plant operation - failure to manually insert rods during an emergency
Nuclear power plant operation - restoring residual heat removal cooling within 10 minutes during an emergency
Nuclear plant - for controls in a well defined mimic layout: select and operate the wrong control during an emergency event
Nuclear power plant operation - actuate the manual depressurisation system during an emergency
Nuclear plant - for controls which are functionally grouped: select and operate the wrong control during an emergency event
Nuclear power plant operation - operate the nuclear instrumentation system during an emergency
Drilling for oil/gas offshore - carry out emergency control action too fast
Nuclear power plant operation - manually operate the reactor core isolation cooling system during an emergency
Drilling for oil/gas offshore - operator fails to isolate drilling equipment but failure is subsequently identified by other rig personnel
Nuclear plant - select and operate the wrong control during an emergency event
Nuclear plant - Incorrect setting of a multi-position selector switch during an emergency event
Drilling for oil/gas offshore - fail to control emergency event, leading to a leak of gas/oil
Nuclear power plant operation - initiate residual heat removal service water (RHRSW) system after a high pressure pool temperature alarm
Nuclear plant - for controls with labels only: select and operate the wrong control during an emergency event
Nuclear plant - Incorrect operation of a manual control during an emergency event
Drilling for oil/gas offshore - calculation incorrectly performed during an emergency
Drilling for oil/gas offshore - too much pressure in system when circulating fluid
Nuclear power plant operation - Taking appropriate action to prevent potential high pressure coolent injection system failure during an emergency.
Nuclear plant - turn control in wrong direction (when the control direction is the same as typical operator expectation) during an emergency event
Nuclear plant - operate a control incorrectly during an emergency
Omission errors during emergency events
Omission of an emergency event task step
Nuclear power plant operation - open valves within one hour in an emergency
Drilling for oil/gas offshore - fail to control system pressure during an emergency
Nuclear plant - fail to recover from error of turning control in wrong direction (when the control direction is the same as typical operator expectation) during an emergency event
Nuclear plant - failure to recover from incorrectly operating a control during an emergency
Omission of an emergency event task (includes events with and without written procedures)
Nuclear power plant operation - Ventilation to cool vital system equipment during an emergency.
Omission of an emergency event task when procedures are not available
Drilling for oil/gas offshore - choke control opened too much
Commission errors during emergency events
Extraneous errors during emergency events
Nuclear power plant operation - Realization of Potential high pressure coolent injection system failure during an emergency.

Drilling for oil/gas offshore - failure to maintain constant pressure due to mis-setting a control
Drilling for oil/gas offshore - finish emergency task too early
Diagnosis and response using hardcopy procedures
All errors during emergency operation using computer-based procedures
Nuclear plant - Failure to recover from the incorrect operation of a manual control during an emergency event
All errors during emergency operation using hard copy, paper-based procedures
Nuclear plant - Fail to recover from error of incorrectly setting of a multi-position selector switch during an emergency event
Drilling for oil/gas offshore - selects incorrect process variable to record
Nuclear plant - turn control in wrong direction (when the correct control direction is opposite to typical operator expectation) during an emergency event
Diagnosis and response using hardcopy or computerised procedures
Nuclear plant - for functionally grouped controls: failure to recover from selecting and operating the wrong control during an emergency event
Diagnosis and response using computerised procedures
Transient 1: Failure in complex diagnosis and response in a central control room using the DISKET decision support tool
Nuclear plant - failure to recover from selecting and operating the wrong control during an emergency event
Transient 2: Failure in complex diagnosis and response in a central control room using the DISKET decision support tool
Nuclear plant - for controls differentiated by text labels only: failure to recover from selecting and operating the wrong control during an emergency event
Transient 1: complex diagnosis and response using a standard NPP central control room interface
Transient 2: Failure in complex diagnosis and response using a standard NPP central control room interface

Acknowledgements

The authors gratefully acknowledge the assistance of Dr Huw Gibson, RSSB, in providing access to COREDATA. Any use of COREDATA in this report is the responsibility of the authors.

References

Cardosi, K. M. 1994, *An Analysis of Tower (Local) Controller-Pilot Voice Communications*, U.S. Department of Transportation, Washington D.C., DOT/FAA/RD-94/11.

Comer, M. K., Seaver, D. A., Stillwell, W. G., & Gaddy, C. D. 1984, *General Human Reliability Estimates Using Expert Judgment, vol 1: Main report*, U.S. Nuclear Regulatory Commission, NUREG/CR-3688.

Cullen, L., Smith, S., Atkins, & IOE 2004, *Rail-specific human reliability assessment technique for driving tasks. Final report*, RSSB.

Dickinson, S. & Lowe, E. 2007, "Supervision in signal boxes," in *People and rail systems - Human factors at the heart of the railway*, J. R. Wilson et al., eds., Aldershot, pp. 225-235.

Gibson, H., Hickling, B., & Kirwan, B. 2006, *Feasibility study into the collection of human error probability data*, EUROCONTROL, Brétigny-sur-Orge Cedex
FRANCE, **EEC Note No. 02/06**.

Hickling, N. 2007, *An Independent Review of a Rail-specific Human Reliability Assessment Technique for Driving Tasks*, RSSB.

Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G., & Umbers, I. 2004, "Nuclear action reliability assessment (NARA): A data-based HRA tool", *PROBABILISTIC SAFETY ASSESSMENT AND MANAGEMENT, VOL 1-6* pp. 1206-1211.

Kirwan, B. 1994, *A guide to practical human reliability assessment* Taylor & Francis.

Kirwan, B. 2008, "Human Reliability Assessment," in *The Encyclopedia of Quantitative Risk Analysis and Assessment*, pp. 1-22.

Kirwan, B. & Gibson, H. 2007, "CARA: A human reliability assessment tool for air traffic safety management - Technical basis and preliminary architecture", *SAFETY OF SYSTEMS* pp. 197-214.

Lee, S. W., Kim, A. R., Ha, J. S., & Seong, P. H. 2011, "Development of a qualitative evaluation framework for performance shaping factors (PSFs) in advanced MCR HRA", *Annals of Nuclear Energy*, vol. 38, no. 8, pp. 1751-1759.

Morrow, D., Lee, A., & Rodvold, M. 1993, "Analysis of Problems in Routine Controller-Pilot Communication", *The International Journal of Aviation Psychology*, vol. 3, no. 4, pp. 285-302.

Rasmussen, J. 1986, *Information Processing and Human-Machine Interaction* North-Holland.

Weber, J. 1992, "The front-line supervisor's role in safety", *Professional Safety*, vol. 37, no. 5.

Williams, J. C. "HEART - a proposed method for assessing and reducing human error", 9th conference on advances in reliability technology.

This report describes an HRA (Human Reliability Assessment) of six generic tasks and four Performance Shaping Factors (PSFs) targeted at railway operations commissioned by Banedanmark. The selection and characterization of generic tasks and PSFs are elaborated by DTU Management in close collaboration with Banedanmark. The estimates provided are based on HRA literature and primarily the HEART method, being recently adapted for railway tasks by the British Rail Safety and Standards Board (RSSB). The method presented in this report differs from the RSSB tool by supporting an analysis at task level, which can be performed with fewer resources than a more detailed analysis of specific errors for each task. The generic tasks are presented with estimated Human Error Probabilities (HEPs) based on and extrapolated from the HRA literature, and estimates are compared with samples of measures from comparable tasks from the COREDATA database. PSFs are presented with multipliers to be used in combination with generic tasks types to support a quantitative HRA of railway tasks. Estimates contained in this report should be used with caution and judgement, since they are largely based on estimates derived from industries other than rail and the general warning that a task-based analysis is less precise than an error-based one. The authors recommend that estimates be adjusted to actual measures of task failures when feasible.

ISBN: 978-87-92706-40-9

DTU Management Engineering
Department of Management Engineering
Technical University of Denmark

Produktionstorvet
Building 424
DK-2800 Kongens Lyngby
Denmark
Tel. +45 45 25 48 00
Fax +45 45 93 34 35

www.man.dtu.dk